Identifiability of Restricted Structural Equation Models

Networks: Processes and Causality, Menorca

Jonas Peters[1,2]

J. Mooij[3], D. Janzing[2], B. Schölkopf[2], R. Tanase[1], P. Bühlmann[1]

[1] Seminar for Statistics, ETH Zürich, Switzerland
[2] MPI for Intelligent Systems, Tübingen, Germany
[3] Radboud University, Nijmegen, Netherlands

3rd September 2012

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

MAX-PLANCK-GESELLSCHAFT

What is the Problem?

Random variables:

X : water temperature of Mediterranean Sea

Y : # networks and causality related workshops in Cala Galdana

Z : # scientists on Menorca

What is the causal structure?

What is the Problem?

Random variables:

X : water temperature of Mediterranean Sea

Y : # networks and causality related workshops in Cala Galdana

Z : # scientists on Menorca

What is the causal structure?

- Understand the (physical) process in more detail.

What is the Problem?

Random variables:

X : water temperature of Mediterranean Sea

Y : # networks and causality related workshops in Cala Galdana

Z : # scientists on Menorca

What is the causal structure?

- Understand the (physical) process in more detail.
- Intervene: Organize workshop in Cala Galdana! Go swimming!

What is the Problem?

Random variables:

X : water temperature of Mediterranean Sea

Y : # networks and causality related workshops in Cala Galdana

Z : # scientists on Menorca

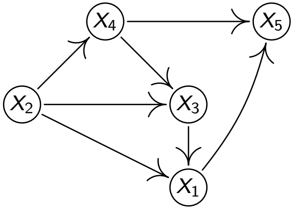
What is the causal structure?

- Understand the (physical) process in more detail.
- Intervene: Organize workshop in Cala Galdana! Go swimming!
- Use observational data!

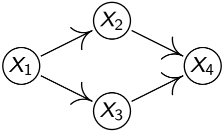
What is the Problem?

observed iid data
from $P(X_1, \ldots, X_5)$ $\xrightarrow{?}$ causal DAG $\mathcal{G}_0$

X_1	3.4	1.7	−2.4	…
X_2	−0.2	7.0	−1.2	…
X_3	−0.1	4.3	−0.7	…
X_4	0.3	5.8	0.3	…
X_5	3.5	1.9	−1.9	…



Relating Causal Graph and Joint Distribution



1. Markov Condition:

$$X_1 \perp\!\!\!\perp X_4 \mid \{X_2, X_3\}$$
$$X_2 \perp\!\!\!\perp X_3 \mid \{X_1\}$$

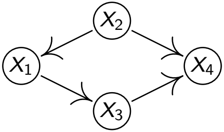
(d-separation $\Rightarrow$ cond. independence)

2. Faithfulness:

no more

(no d-separation $\Rightarrow$ no cond. independence)

Relating Causal Graph and Joint Distribution



1. Markov Condition:

$$X_1 \perp\!\!\!\perp X_4 \mid \{X_2, X_3\}$$
$$X_2 \perp\!\!\!\perp X_3 \mid \{X_1\}$$

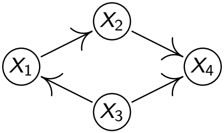
(d-separation $\Rightarrow$ cond. independence)

2. Faithfulness:

no more

(no d-separation $\Rightarrow$ no cond. independence)

Relating Causal Graph and Joint Distribution



1. Markov Condition:

$$X_1 \perp\!\!\!\perp X_4 \mid \{X_2, X_3\}$$
$$X_2 \perp\!\!\!\perp X_3 \mid \{X_1\}$$

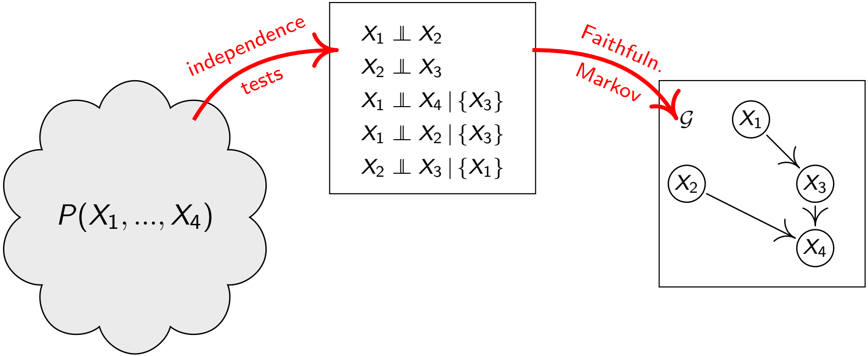
(d-separation $\Rightarrow$ cond. independence)

2. Faithfulness:

no more

(no d-separation $\Rightarrow$ no cond. independence)

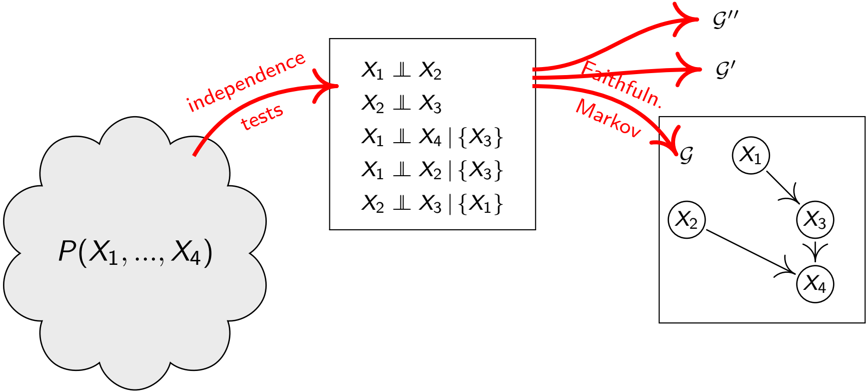
PC Algorithm



Method: PC [Spirtes et al., 2001]

1. Find all (cond.) independences from the data.
2. Select the DAG(s) that corresponds to these independences.

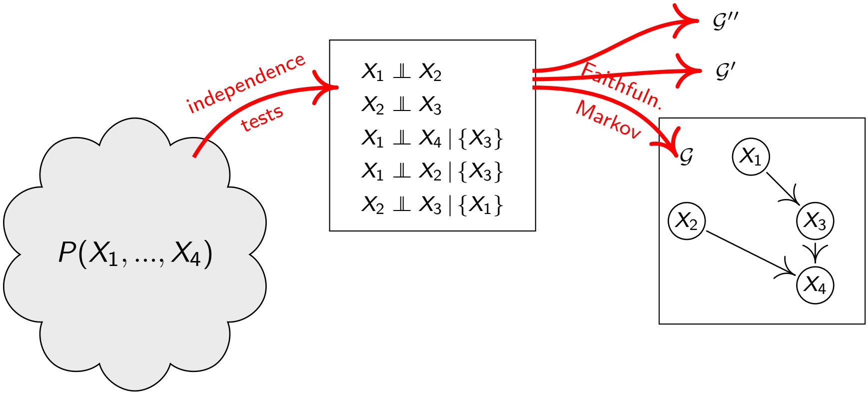
PC Algorithm



Method: PC [Spirtes et al., 2001]

1. Find all (cond.) independences from the data.
2. Select the DAG(s) that corresponds to these independences.

PC Algorithm



Method: PC [Spirtes et al., 2001]

1. ~~Find all (cond.) independences from the data.~~ Be smart.
2. Select the DAG(s) that corresponds to these independences.

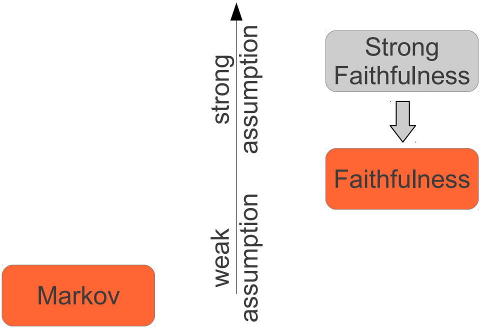
Relating Causal Graph and Joint Distribution

The PC algorithm makes very few assumptions.

Can we gain something by making more/different assumptions?

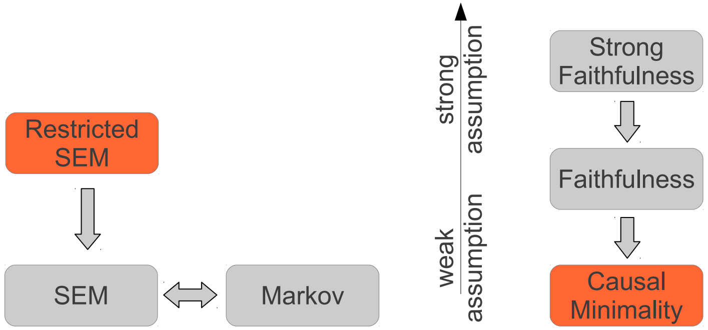
Relating Causal Graph and Joint Distribution

PC assumptions:



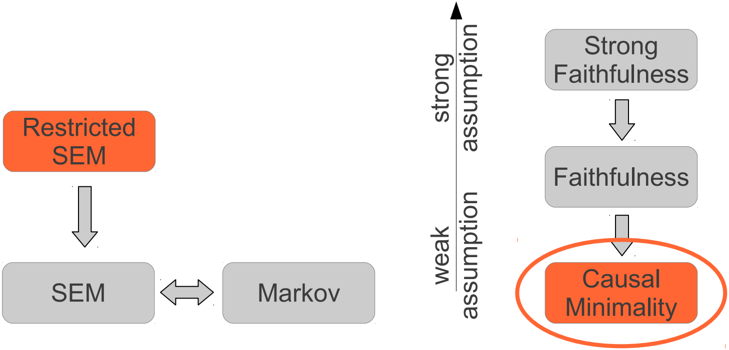
Relating Causal Graph and Joint Distribution

New assumptions:



Relating Causal Graph and Joint Distribution

New assumptions:



Causal Minimality

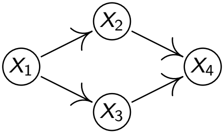
Causal Minimality is a weak form of faithfulness:

Definition

Let $\mathcal{G}_0$ be the true causal graph. If $P(X_1, \ldots, X_p)$ is not Markov to any proper subgraph of $\mathcal{G}_0$, causal minimality is satisfied.

$\rightsquigarrow$ $\boxed{\text{“Each arrow does something.”}}$

Violation of Causal Minimality



1. Markov Condition:

$$X_2 \perp\!\!\!\perp X_3 \mid \{X_1\}$$
$$X_1 \perp\!\!\!\perp X_4 \mid \{X_2, X_3\}$$

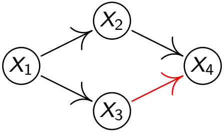
(d-separation $\Rightarrow$ cond. independence)

2. Faithfulness:

no more

(no d-separation $\Rightarrow$ no cond. independence)

Violation of Causal Minimality



1. Markov Condition:

$$X_2 \perp\!\!\!\perp X_3 \,|\, \{X_1\}$$
$$X_1 \perp\!\!\!\perp X_4 \,|\, \{X_2, X_3\}$$
$$X_4 \perp\!\!\!\perp X_3 \,|\, \{X_2\}$$

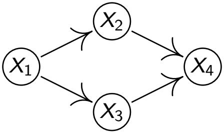
(d-separation $\Rightarrow$ cond. independence)

2. ~~Faithfulness:~~

~~no more~~

(no d-separation $\Rightarrow$ no cond. independence)

Violation of Faithfulness



1. Markov Condition:

$X_2 \perp\!\!\!\perp X_3 \mid \{X_1\}$
$X_1 \perp\!\!\!\perp X_4 \mid \{X_2, X_3\}$
$X_1 \perp\!\!\!\perp X_4$

(d-separation $\Rightarrow$ cond. independence)

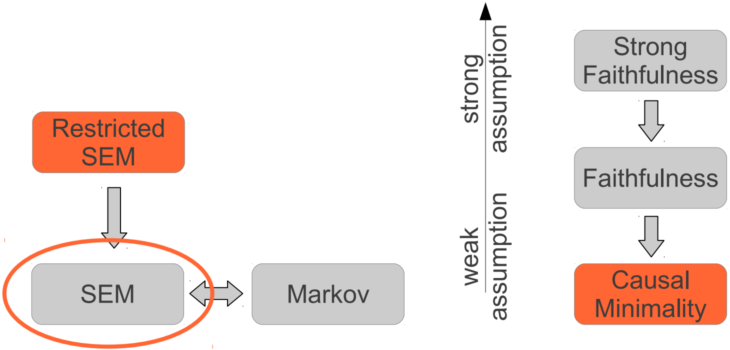
2. ~~Faithfulness:~~

~~no more~~

(no d-separation $\Rightarrow$ no cond. independence)

Relating Causal Graph and Joint Distribution

New assumptions:



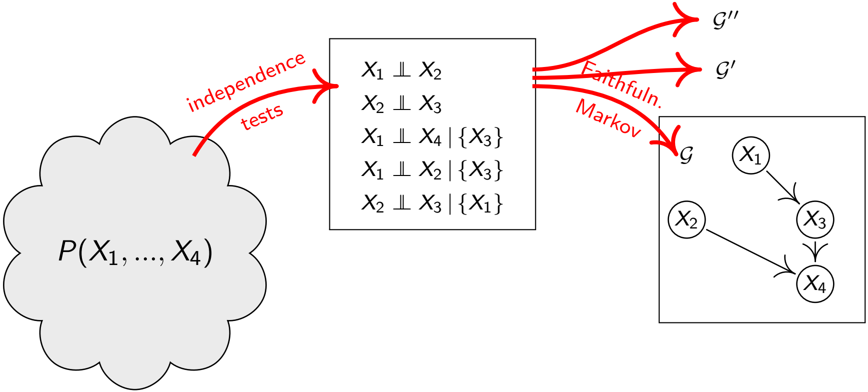
Structural Equation Models

The joint distribution $P(X_1, \ldots, X_p)$ satisfies a Structural Equation Model (SEM) with graph $\mathcal{G}_0$ if

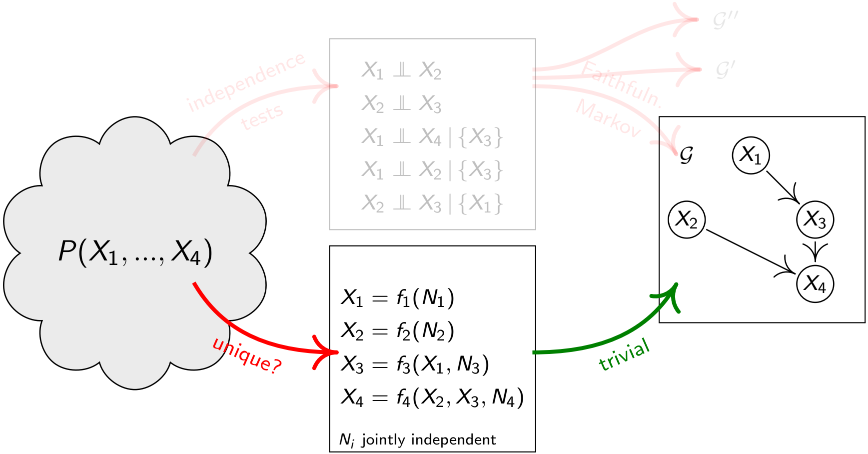
$$X_i = f_i(\mathbf{PA}_i, N_i) \qquad 1 \le i \le p$$

with $\mathbf{PA}_i$ being the parents of X_i in $\mathcal{G}_0$. The N_i are required to be jointly independent.

The Alternative Route

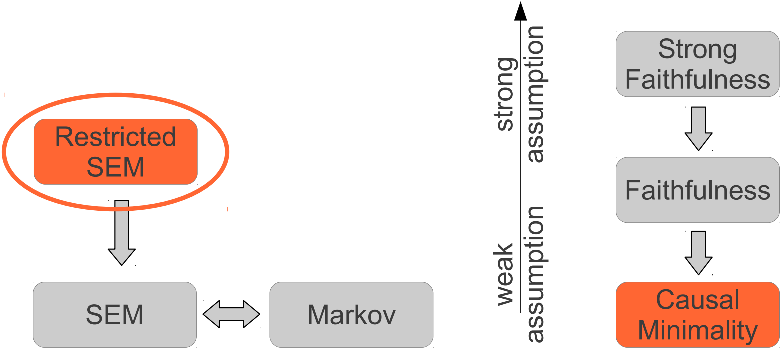


The Alternative Route



Relating Causal Graph and Joint Distribution

New assumptions:



Restricted Structural Equation Models

- Linear Gaussian Additive Noise Models

$$X_i = \sum_{j \in \mathbf{PA}_i} \beta_j X_j + N_i \qquad 1 \leq i \leq p$$

with $N_i \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_i^2)$ and graph $\mathcal{G}_0$.

Proposition

Assume faithfulness. Then one can identify the Markov equivalence class of $\mathcal{G}_0$ from $P(X_1, \ldots, X_p)$.

Restricted Structural Equation Models

- Linear Non-Gaussian Additive Noise Models

$$X_i = \sum_{j \in \mathbf{PA}_i} \beta_j X_j + N_i \qquad 1 \leq i \leq p$$

with $N_i \overset{\text{iid}}{\sim}$ non-Gaussian and graph $\mathcal{G}_0$.
(One can show: $\beta_j \neq 0 \Rightarrow$ causal minimality.)

Theorem ([Shimizu et al., 2006])

One can identify $\mathcal{G}_0$ from $P(X_1, \ldots, X_p)$.

Restricted Structural Equation Models

- Linear Gaussian Models with same Error Variance

$$X_i = \sum_{j \in \mathbf{PA}_i} \beta_j X_j + N_i \qquad 1 \le i \le p$$

with $N_i \stackrel{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2)$.
(One can show: $\beta_j \neq 0 \Rightarrow$ causal minimality.)

Theorem ([Peters and Bühlmann, 2012])

One can identify $\mathcal{G}_0$ from $P(X_1, \ldots, X_p)$.

Restricted Structural Equation Models

- Non-Linear Additive Noise Models

$$X_i = f_i(X_{\mathbf{PA}_i}) + N_i \qquad 1 \leq i \leq p$$

with N_i iid and graph $\mathcal{G}_0$.

Theorem ([Hoyer et al., 2009, Peters et al., 2011b])

Exclude a few combinations of f_i and N_i. Then one can identify $\mathcal{G}_0$ from $P(X_1, \ldots, X_p)$.

Restricted Structural Equation Models

- Discrete Additive Noise Models

$$X_i = f_i(X_{\mathbf{PA}_i}) + N_i \qquad 1 \le i \le p$$

with $N_i \overset{\text{iid}}{\sim}$ non-uniform and graph $\mathcal{G}_0$.

Theorem ([Peters et al., 2011a,b])

Exclude a few combinations of f_i and N_i. Then one can identify $\mathcal{G}_0$ from $P(X_1, \dots, X_p)$.

Restricted Structural Equation Models

Assumption

Assume that $P(X_1, \ldots, X_p)$ follows any of the restricted SEMs mentioned above with graph $\mathcal{G}_0$ and assume causal minimality.

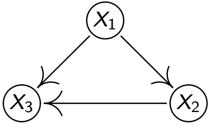
Theorem

Then, the true causal DAG can be recovered from the joint distribution.

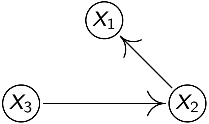
Linear Gaussian Models with fixed Variance

Proof Idea:

Assume $P(X_1, X_2, X_3)$ allows for two SEMs leading to $\mathcal{G}_1$ and $\mathcal{G}_2$:



$\mathcal{G}_1$



$\mathcal{G}_2$

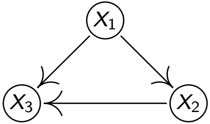
$X_3 = \alpha_1 X_1 + \alpha_2 X_2 + N_3$

$X_3 = M_3$

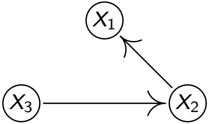
Linear Gaussian Models with fixed Variance

Proof Idea:

Assume $P(X_1, X_2, X_3)$ allows for two SEMs leading to $\mathcal{G}_1$ and $\mathcal{G}_2$:



$\mathcal{G}_1$



$\mathcal{G}_2$

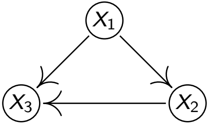
$$X_3 = \alpha_1 X_1 + \alpha_2 X_2 + N_3$$
$$X_3^* := X_3|_{X_1=x} = \alpha_1 x + \alpha_2 X_2|_{X_1=x} + N_3$$

$$X_3 = M_3$$
$$X_3^* := X_3|_{X_1=x} = M_3|_{X_1=x}$$

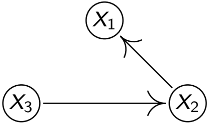
Linear Gaussian Models with fixed Variance

Proof Idea:

Assume $P(X_1, X_2, X_3)$ allows for two SEMs leading to $\mathcal{G}_1$ and $\mathcal{G}_2$:



$\mathcal{G}_1$



$\mathcal{G}_2$

$$
\begin{aligned}
X_3 &= \alpha_1 X_1 + \alpha_2 X_2 + N_3 \\
X_3^* := X_3|_{X_1=x} &= \alpha_1 x + \alpha_2 X_2|_{X_1=x} + N_3 \\
\Rightarrow \quad \mathrm{var}(X_3^*) &= 0 + \alpha_2^2 \mathrm{var}(X_2|_{X_1=x}) + \sigma^2 > \sigma^2
\end{aligned}
$$

$$
\begin{aligned}
X_3 &= M_3 \\
X_3^* := X_3|_{X_1=x} &= M_3|_{X_1=x} \\
\Rightarrow \quad \mathrm{var}(X_3^*) &\leq \sigma^2
\end{aligned}
$$

Practical Method I

Method: IFMOC (Identifiable Functional Model Class)

Idea: If we fit a wrong SEM, noise variables become dependent.

1. Find *all* SEMs that fit the data.
2. If there is exactly one, output the DAG. Otherwise: “I do not know”.
3. Avoid enumerating all DAGs [Mooij et al., 2009]: always find sink and remove additional edges at the end.

needed:

- regression method (e.g. linear, GP),
- independence test (e.g. HSIC).

Practical Method II

Method: GDS (Greedy DAG Search)

Only for: linear Gaussian models with same noise variances.

Idea: Define a score (e.g. BIC) to a given DAG.

1. Start with random DAG.
2. At each step, look at all neighbouring DAGs.
3. Go to DAG with best score.

Practical Method - Properties

Both:

- $+$ Identifiability within Markov equivalence class.
- $+$ Option to say “I do not know.”
- $+$ No faithfulness.
- \- Strong structural assumptions.
- \- Not scalable to high-dimensional problems (yet :-)).

Restricted Structural Equation Models

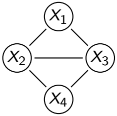
Experiment 1: Comparison IFMOC and PC when both assumptions are met.

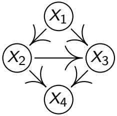
sample size: 400
#data sets: 100
$\alpha = 5\%$

$$X_1 = N_1$$
$$X_2 = N_2$$
$$X_3 = f_3(X_1, X_2) + N_3$$
$$X_4 = f_4(X_2, X_3) + N_4$$

$N_i \overset{\text{iid}}{\sim} \mathcal{U}([-0.5, 0.5])$.

	linear	nonlinear
$\mathrm{PC}_{\mathrm{lin}}$	47%, 53%	100%
$\mathrm{PC}_{\mathrm{nonlin}}$	3%, 97%	4%, 96%
$\mathrm{IFMOC}_{\mathrm{lin}}$	14%, 86%	100%
$\mathrm{IFMOC}_{\mathrm{nonlin}}$	23%, 1%, 76%	6%, 8%, 86%





correct/wrong/no decision

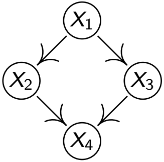
Restricted Structural Equation Models

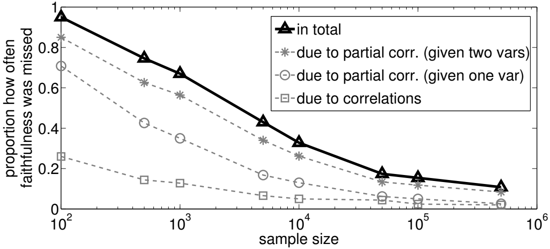
Experiment 2: How often are we close to non-faithfulness?

#data sets: 500, $\alpha = 5\%$.

$$
\begin{aligned}
X_1 &= \beta_1 N_1 \\
X_2 &= \gamma_{12} X_1 + \beta_2 N_2 \\
X_3 &= \gamma_{13} X_1 + \beta_3 N_3 \\
X_4 &= \gamma_{24} X_2 + \gamma_{34} X_3 + \beta_4 N_4
\end{aligned}
$$

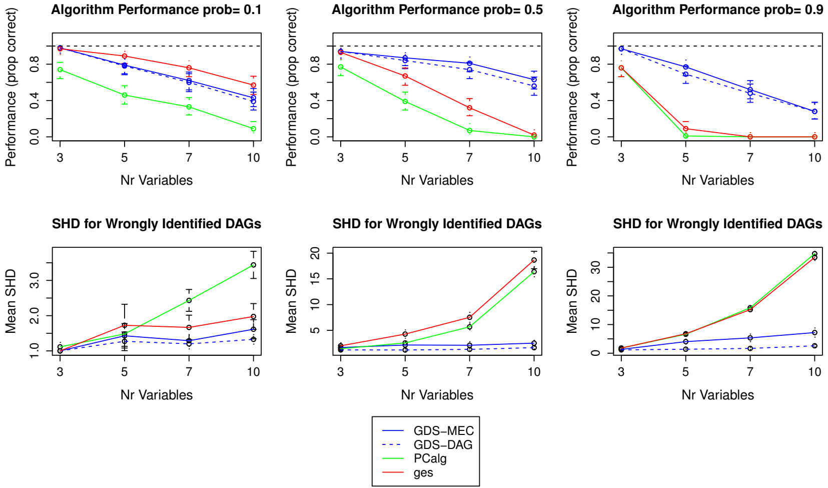
$N_i \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$,
$\gamma_{ij} \overset{\text{iid}}{\sim} \mathcal{U}([-5,5])$, $\beta_i \overset{\text{iid}}{\sim} \mathcal{U}([0,0.5])$.





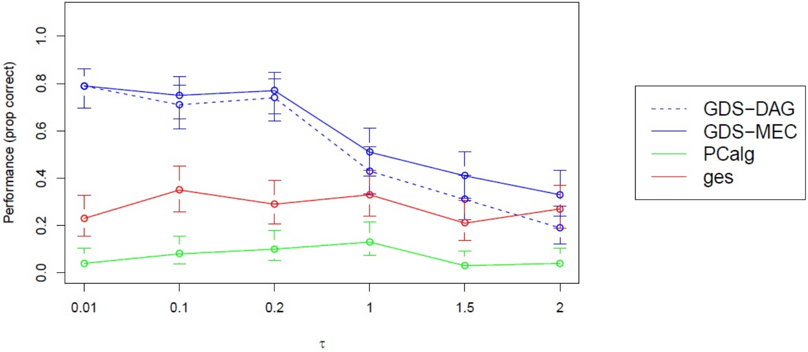
Restricted Structural Equation Models

Experiment 3: Linear Gaussian Models with Fixed Variances (GDS).



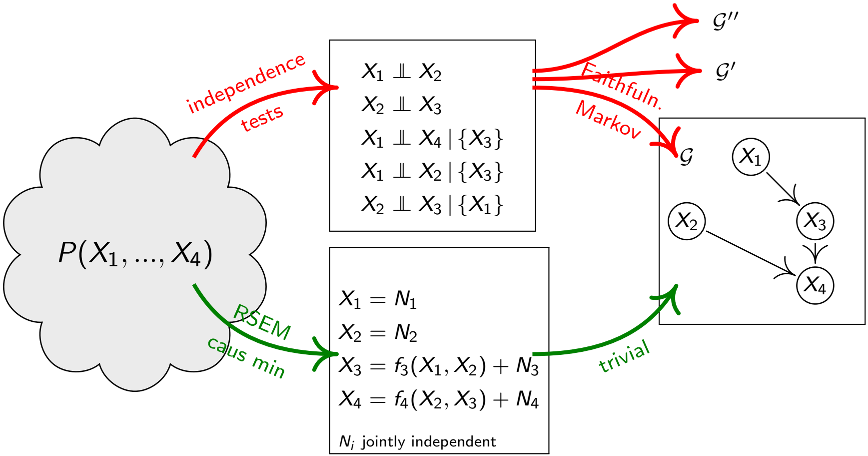
coefs sampled from $\mathcal{U}([-1.5, -0.1] \cup [0.1, 1.5])$; $n = 1000$; 100 repetitions

Experiment 4: Violation of Same Error Variances.



noise variances sampled from $\mathcal{U}([4-\tau, 4+\tau])$
$n = 1000$, $p = 7$, $prob = 0.5$, 100 repetitions

Summary



¡Muchas gracias!

References I

****

P. Hoyer, D. Janzing, J. Mooij, J. Peters, and B. Schölkopf. Nonlinear causal discovery with additive noise models. In *NIPS 21*, 2009.

D. Janzing and B. Schölkopf. Causal inference using the algorithmic Markov condition. *IEEE Trans. on Information Theory*, 56 (10):5168–5194, 2010.

D. Janzing and B. Steudel. Justifying additive-noise-model based causal discovery via algorithmic information theory. *Open Systems and Information Dynamics*, 17:189–212, 2010.

J. Lemeire and E. Dirkx. Causal models as minimal descriptions of multivariate systems. `http://parallel.vub.ac.be/~jan/`, 2006.

C. Meek. Causal inference and causal explanation with background knowledge. In *UAI 11*, pages 403–441, 1995.

J. Mooij, D. Janzing, J. Peters, and B. Schölkopf. Regression by dependence minimization and its application to causal inference. In *ICML 26*, 2009.

J. Peters and P. Bühlmann. Identifiability of Gaussian Structural Equation Models with Same Error Variances. *ArXiv e-prints*, 2012.

J. Peters, D. Janzing, and B. Schölkopf. Causal inference on discrete data using additive noise models. *IEEE Trans. Pattern Analysis Machine Intelligence*, 33(12):2436–2450, 2011a.

J. Peters, J. M. Mooij, D. Janzing, and B. Schölkopf. Identifiability of causal graphs using functional models. In *UAI 27*, 2011b.

S. Shimizu, P. O. Hoyer, A. Hyvärinen, and A. J. Kerminen. A linear non-Gaussian acyclic model for causal discovery. *Journal of Machine Learning Research*, 7:2003–2030, 2006.

P. Spirtes, C. Glymour, and R. Scheines. *Causation, Prediction, and Search*. MIT Press, 2nd edition, 2001.

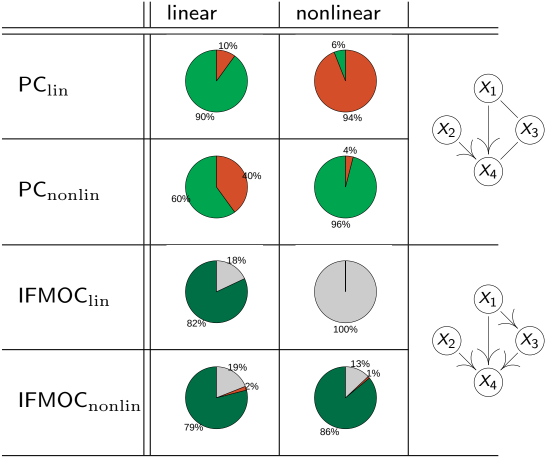
K. Zhang and A. Hyvärinen. On the identifiability of the post-nonlinear causal model. In *UAI 25*, 2009.

Experiment 1a: Both methods should work when both assumptions are met.

sample size: 400
#data sets: 100
$\alpha = 5\%$

$$X_1 = N_1$$
$$X_2 = N_2$$
$$X_3 = f_3(X_1) + N_3$$
$$X_4 = f_4(X_1, X_2, X_3) + N_4$$

$$N_i \overset{\text{iid}}{\sim} \mathcal{U}([-0.5, 0.5]).$$



correct/wrong/no decision

Future Work

- Understand relations to Bayesian Network Learning.
- Joint independence of noise $\leftrightarrow$ joint independence noise to ancestors.
- Discrete Confounder.
- Extensive tests on real data, especially on data sets with > 2 variables.
- Robustness.

Markov Condition and Faithfulness

Let $\mathcal{G}$ be the true causal graph of $X_1, \ldots, X_p$.

Assumption (Markov Assumption)

$$X_i \text{ and } X_j \text{ are d-separated by } \mathcal{S} \text{ in } \mathcal{G} \quad \Rightarrow \quad X_i \perp\!\!\!\perp X_j \,|\, \mathcal{S}$$

Markov Condition and Faithfulness

Let $\mathcal{G}$ be the true causal graph of $X_1, \ldots, X_p$.

Assumption (Markov Assumption)

$$X_i \textit{ and } X_j \textit{ are d-separated by } \mathcal{S} \textit{ in } \mathcal{G} \quad \Rightarrow \quad X_i \perp\!\!\!\perp X_j \,|\, \mathcal{S}$$

Assumption (Faithfulness Assumption)

$$X_i \textit{ and } X_j \textit{ are d-separated by } \mathcal{S} \textit{ in } \mathcal{G} \quad \Leftarrow \quad X_i \perp\!\!\!\perp X_j \,|\, \mathcal{S}$$

Method: IFMOC (two variables)

1. Assume an ANM from cause to effect.
2. Fit $Y = f(X) + N$ and $X = g(Y) + M$ and check which of the two models lead to independent residuals.
3. If only one direction does, output it. Otherwise do not decide.

Independence of Conditional and Marginal

Suppose X is the cause and Y effect. What if

$$Y \neq f(X) + N, \quad N \perp\!\!\!\perp X\,,$$

but

$$X = g(Y) + M, \quad M \perp\!\!\!\perp Y\,?$$

Independence of Conditional and Marginal

Suppose X is the cause and Y effect. What if

$$Y \neq f(X) + N, \quad N \perp\!\!\!\perp X\,,$$

but

$$X = g(Y) + M, \quad M \perp\!\!\!\perp Y\,?$$

Janzing and Steudel [2010]: This implies "dependence" (based on Kolmogorov complexity) between

$$p(\text{cause}) \text{ and } p(\text{effect} \,|\, \text{cause})$$

One rather expects input and mechanism to be most often "independent" [Lemeire and Dirkx, 2006, Janzing and Schölkopf, 2010].

Identifiable Functional Model Class (IFMOC)

Definition (Bivariate Identifiable Set)

We call a set $\mathcal{B} \subseteq \mathcal{F} \times \mathcal{P}_{\mathbb{R}} \times \mathcal{P}_{\mathbb{R}}$ containing combinations of functions $f \in \mathcal{F}$ and distributions $P(X)$, $P(N)$ of input X and noise N *bivariate identifiable in* $\mathcal{F}$ if the following holds:

Identifiable Functional Model Class (IFMOC)

Definition (Bivariate Identifiable Set)

We call a set $\mathcal{B} \subseteq \mathcal{F} \times \mathcal{P}_{\mathbb{R}} \times \mathcal{P}_{\mathbb{R}}$ containing combinations of functions $f \in \mathcal{F}$ and distributions $P(X)$, $P(N)$ of input X and noise N *bivariate identifiable in* $\mathcal{F}$ if the following holds:

$$
\begin{array}{cccc}
 & (f, P(X), P(N)) \in \mathcal{B} & \text{and} & Y = f(X, N), N \perp\!\!\!\perp X \\
\Rightarrow & \nexists g \in \mathcal{F}: & & X = g(Y, M), M \perp\!\!\!\perp Y
\end{array}
$$

Identifiable Functional Model Class (IFMOC)

Definition (Bivariate Identifiable Set)

We call a set $\mathcal{B} \subseteq \mathcal{F} \times \mathcal{P}_{\mathbb{R}} \times \mathcal{P}_{\mathbb{R}}$ containing combinations of functions $f \in \mathcal{F}$ and distributions $P(X)$, $P(N)$ of input X and noise N *bivariate identifiable in* $\mathcal{F}$ if the following holds:

$$\begin{array}{cccc} & (f, P(X), P(N)) \in \mathcal{B} & \text{and} & Y = f(X,N), N \perp\!\!\!\perp X \\ \Rightarrow & \nexists g \in \mathcal{F}: & & X = g(Y,M), M \perp\!\!\!\perp Y \end{array}$$

Additionally we require

$$f(X,N) \not\perp\!\!\!\perp X \tag{1}$$

for all $(f, P(X), P(N)) \in \mathcal{B}$ with $N \perp\!\!\!\perp X$.

Identifiable Functional Model Class (IFMOC)

Lemma

The following sets are bivariate identifiable:

(i) *linear ANMs [Shimizu et al., 2006]:* $\mathcal{F}_1 = \{f(x,n) = ax + n\}$

$$\mathcal{B}_1 = \{(X,N) \text{ not both Gaussian}\} \setminus \tilde{B}_1$$

(ii) *discrete ANMs [Peters et al., 2011b]:* $\mathcal{F}_2 = \{f(x,n) \equiv \phi(x) + n(\tilde{m})\}$

$$\mathcal{B}_2 = \{(\phi, X) \text{ not affine and uniform}\} \setminus \tilde{B}_2$$

(iii) *non-linear ANMs [Hoyer et al., 2009]*

$$\mathcal{B}_3 = \{(\phi, X, N) \text{ not lin., Gauss, Gauss}\} \setminus \tilde{B}_3$$

(iv) *post-nonlinear ANMs [Zhang and Hyvärinen, 2009]*

Identifiable Functional Model Class (IFMOC)

How can we transfer these identifiability results to p variables?

Identifiable Functional Model Class (IFMOC)

How can we transfer these identifiability results to p variables?

Definition ($\mathcal{F}$-FMOC)

- p equations

$$X_i = f_i(\mathbf{PA}_i, N_i) \qquad 1 \leq i \leq p$$

are called a *functional model* if N_i are jointly independent and the corresponding graph is acyclic.

Identifiable Functional Model Class (IFMOC)

How can we transfer these identifiability results to p variables?

Definition ($\mathcal{F}$-FMOC)

- p equations
$$X_i = f_i(\mathbf{PA}_i, N_i) \qquad 1 \leq i \leq p$$
are called a *functional model* if N_i are jointly independent and the corresponding graph is acyclic.
- A set of functional models is called a *functional model class with function class $\mathcal{F}$, for short $\mathcal{F}$-FMOC*, if each of the functional models satisfies $f_i \in \mathcal{F}$ for all i.

Identifiable Functional Model Class (IFMOC)

Definition ($(\mathcal{B}, \mathcal{F})$-IFMOC)

Let $\mathcal{B}$ be bivariate identifiable in $\mathcal{F}$. An $\mathcal{F}$-FMOC is called a $(\mathcal{B}, \mathcal{F})$-*Identifiable Functional Model Class*, for short $(\mathcal{B}, \mathcal{F})$-IFMOC, if for all its functional models

$$X_i = f_i(\mathbf{PA}_i, N_i)\,, \qquad 1 \leq i \leq p$$

and for all $1 \leq i \leq p$, $j \in \mathbf{PA}_i$, for all sets $\mathbf{S} \subseteq \{1, \dots, p\}$ with $\mathbf{PA}_i \setminus \{j\} \subseteq \mathbf{S} \subseteq \mathbf{ND}_i \setminus \{i, j\}$ we have:
There exists an $x_\mathbf{S}$ with $p_\mathbf{S}(x_\mathbf{S}) > 0$ and

$$\Big(f_i(x_{\mathbf{PA}_i \setminus \{j\}}, \underbrace{\cdot}_{X_j}, \underbrace{\cdot}_{N_i}), P(X_j \,|\, X_\mathbf{S} = x_\mathbf{S}), P(N_i)\Big) \in \mathcal{B}\,. \qquad (2)$$

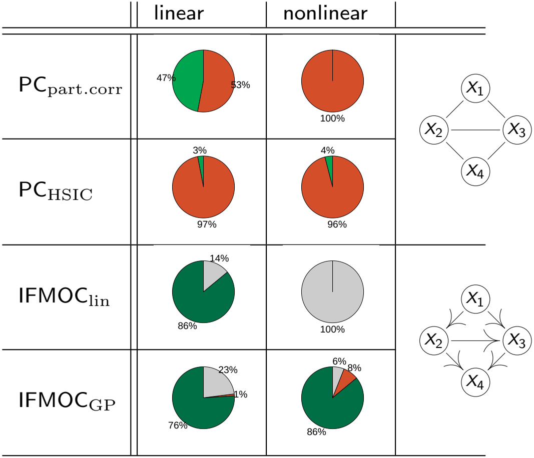
Identifiable Functional Model Class (IFMOC)

Experiment 1b: Both methods should work when both assumptions are met.

sample size: 400
$\#$data sets: 100
$\alpha = 5\%$

$$X_1 = f_1(N_1)$$
$$X_2 = f_2(N_2)$$
$$X_3 = f_3(X_1, X_2, N_3)$$
$$X_4 = f_4(X_2, X_3, N_4)$$

$N_i \stackrel{\text{iid}}{\sim} \mathcal{U}([-0.5, 0.5])$.



correct/wrong/no decision

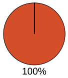
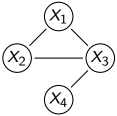
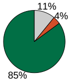
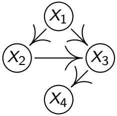
Restricted Structural Equation Models

Experiment 2: If the distribution is not faithful, PC fails, IFMOC does not.

sample size: 1000
#data sets: 100
$\alpha = 5\%$

$$X_1 = N_1$$
$$X_2 = 1.5X_1 + N_2$$
$$X_3 = 3X_1 - 2X_2 + N_3$$
$$X_4 = 1.8X_3 + N_4$$

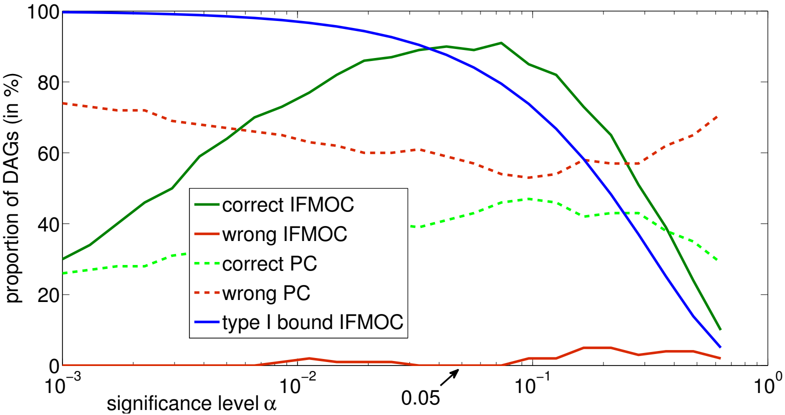
with $N_i \overset{\text{iid}}{\sim} \mathcal{U}([0, 0.5])$.

	linear	
$\mathrm{PC}_{\mathrm{lin}}$		
$\mathrm{IFMOC}_{\mathrm{lin}}$		

correct/wrong/no decision

Identifiable Functional Model Class (IFMOC)

Experiment 2b: Both methods should work when both assumptions are met.



Identifiable Functional Model Class (IFMOC)

already known: 2 variable case

Theorem (Hoyer et al. [2009])

Let

$$Y = f(X) + N, \quad N \perp\!\!\!\perp X\,.$$

Then for most combinations $(f, P(X), P(N))$

$$X \neq g(Y) + M, \quad M \perp\!\!\!\perp Y\,.$$

Those combinations $(f, P(X), P(N))$ *are called bivariate identifying.*

Identifiable Functional Model Class (IFMOC)

already known: 2 variable case

Theorem (Hoyer et al. [2009])

Let

$$Y = f(X) + N, \quad N \perp\!\!\!\perp X\,.$$

Then for most combinations $(f, P(X), P(N))$

$$X \neq g(Y) + M, \quad M \perp\!\!\!\perp Y\,.$$

Those combinations $(f, P(X), P(N))$ *are called bivariate identifying.*

Similar results for

(i) post-nonlinear additive noise [Zhang et al., 2009]

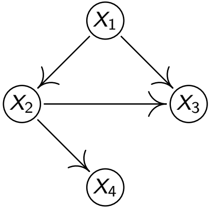
(ii) discrete additive noise [Peters et al., 2011b]

Identifiable Functional Model Class (IFMOC)

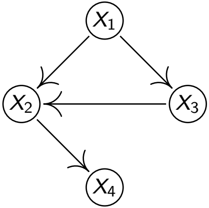
What happens in the case of p variables?

Identifiable Functional Model Class (IFMOC)

Assume $P(X_1, X_2, X_3, X_4)$ allows for two functional models leading to $\mathcal{G}_1$ and $\mathcal{G}_2$:



$\mathcal{G}_1$

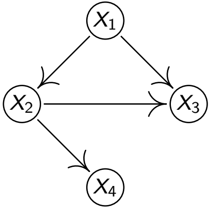


$\mathcal{G}_2$

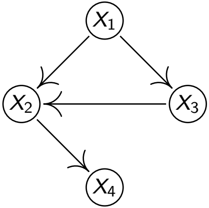
$$X_3 = f(X_1, X_2, N_3) \qquad X_2 = g(X_1, X_3, N_2)$$

Identifiable Functional Model Class (IFMOC)

Assume $P(X_1, X_2, X_3, X_4)$ allows for two functional models leading to $\mathcal{G}_1$ and $\mathcal{G}_2$:



$\mathcal{G}_1$



$\mathcal{G}_2$

$$X_3 = f(X_1, X_2, N_3) \qquad\qquad X_2 = g(X_1, X_3, N_2)$$

$$\Rightarrow \quad X_3|_{X_1=x} = f(x, X_2|_{X_1=x}, N_3)$$

Identifiable Functional Model Class (IFMOC)

Assume $P(X_1, X_2, X_3, X_4)$ allows for two functional models leading to $\mathcal{G}_1$ and $\mathcal{G}_2$:

$$\begin{tikzcd} & X_1 \arrow[dl] \arrow[dr] & \\ X_2 \arrow[rr] \arrow[dr] & & X_3 \\ & X_4 & \end{tikzcd} \qquad \mathcal{G}_1$$

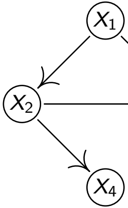
$$\begin{tikzcd} & X_1 \arrow[dl] \arrow[dr] & \\ X_2 \arrow[dr] & & X_3 \arrow[ll] \\ & X_4 & \end{tikzcd} \qquad \mathcal{G}_2$$

$$X_3 = f(X_1, X_2, N_3) \qquad\qquad X_2 = g(X_1, X_3, N_2)$$

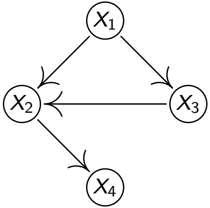
$$\Rightarrow \quad \begin{aligned} X_3|_{X_1=x} &= f(x, X_2|_{X_1=x}, N_3) \\ X_2|_{X_1=x} &= g(x, X_3|_{X_1=x}, N_2) \end{aligned}$$

Identifiable Functional Model Class (IFMOC)

Assume $P(X_1, X_2, X_3, X_4)$ allows for two functional models leading to $\mathcal{G}_1$ and $\mathcal{G}_2$:



$\mathcal{G}_1$



$\mathcal{G}_2$

$$X_3 = f(X_1, X_2, N_3) \qquad\qquad X_2 = g(X_1, X_3, N_2)$$

$$\Rightarrow \quad \begin{aligned} X_3|_{X_1=x} &= f(x, X_2|_{X_1=x}, N_3) \\ X_2|_{X_1=x} &= g(x, X_3|_{X_1=x}, N_2) \end{aligned}$$

If the triple $(f(x, \cdot, \cdot), P(X_2|_{X_1=x}), P(N_3))$ is bivariate identifying, then $\lightning$.

Identifiable Functional Model Class (IFMOC)

Definition (IFMOC)

- A set of functional models is called a *functional model class with function class $\mathcal{F}$, for short $\mathcal{F}$-FMOC*, if each of the functional models satisfies $f_i \in \mathcal{F}$ for all i.
- An $\mathcal{F}$-FMOC is called an Identifiable Functional Model Class, for short IFMOC, if for all its functional models

$$X_i = f_i(\mathbf{PA}_i, N_i)\,, \qquad 1 \leq [illegible]$$

and for all $1 \leq i \leq p$, $j \in \mathbf{PA}_i$, for all sets $\mathbf{S} \subseteq \{1$ [illegible] $\setminus \{j\} \subseteq \mathbf{S} \subseteq \mathbf{ND}_i \setminus \{i, j\}$ there exists an $x_{\mathbf{S}}$ with $p_{\mathbf{S}}(x_{\mathbf{S}}) > 0$ and

$$\Big(f_i(\underbrace{\ldots}_{X_j}, \underbrace{\ldots}_{N_i}), P(X_j \mid X_{\mathbf{S}} = x_{\mathbf{S}}), P(N_i) \Big) \text{ is bivariate identifying .} \qquad (3)$$

based on bivariate identifiability